Check for updates

High compression deep learning based single-pixel hyperspectral macroscopic fluorescence lifetime imaging *in vivo*

M. OCHOA,[1] **A. RUDKOUSKAYA,**[2] **R. YAO,**[1] **P. YAN,**[1] **M. BARROSO,**[1] **AND X. INTES**[1,*]

[1]*Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA*
[2]*Department of Molecular and Cellular Physiology, Albany Medical College, Albany, NY 12208, USA*
[*]*intesx@rpi.edu*

Abstract: Single pixel imaging frameworks facilitate the acquisition of high-dimensional optical data in biological applications with photon starved conditions. However, they are still limited to slow acquisition times and low pixel resolution. Herein, we propose a convolutional neural network for fluorescence lifetime imaging with compressed sensing at high compression (NetFLICS-CR), which enables in vivo applications at enhanced resolution, acquisition and processing speeds, without the need for experimental training datasets. NetFLICS-CR produces intensity and lifetime reconstructions at 128×128 pixel resolution over 16 spectral channels while using only up to 1% of the required measurements, therefore reducing acquisition times from ~2.5 hours at 50% compression to ~3 minutes at 99% compression. Its potential is demonstrated in silico, in vitro and for mice in vivo through the monitoring of receptor-ligand interactions in liver and bladder and further imaging of intracellular delivery of the clinical drug Trastuzumab to HER2-positive breast tumor xenografts. The data acquisition time and resolution improvement through NetFLICS-CR, facilitate the translation of single pixel macroscopic fluorescence lifetime imaging (SP-MFLI) for in vivo monitoring of lifetime properties and drug uptake.

1. Introduction

Over the last two decades, computational imaging has experienced an explosive growth thanks to its aptitude to overcome limitations of conventional imaging methods [1]. Especially, optical computational imaging has led to the development of new imaging techniques that have greatly impacted numerous fields including material sciences, computer vision and biomedical sciences [2]. These technical breakthroughs have been stimulated by the wide availability of new optical components such as light modulators and/or sensitive digital sensors. Moreover, the emergence of the theoretical framework of compressive sensing has empowered the implementation of efficient computational imaging systems [3]. Among all technical approaches, structured light techniques have played a significant role in advancing this field [4,5]. For instance, the incorporation of spatial light modulators in the optical chain has led to improved resolution in microscopy [6], enabled the reconstruction of 4D phase amplitude data [7,8], facilitated hyperspectral microscopy [9,10] or enabled fast optical tomography of thick specimens [11,12]. Nevertheless, there is still a strong incentive to develop novel imaging techniques for biomedical applications, that can acquire high-dimensional data (space, time, phase and spectrum), with increased information content, such as required in highly multiplexed molecular imaging. [13] However, biological applications, especially *in vivo* fluorescence-based imaging applications are characterized by faint signals due to the scattering and absorptive nature of intact (non-cleared) tissues [14]. One computational imaging approach that is well positioned to tackle such low signal conditions to generate these desired high-dimensional datasets is single-pixel imaging [15]. To date, the single pixel imaging approach has been implemented for various applications ranging from astronomy

to medical imaging. [16] A single-pixel camera is typically based on an inverse problem that aims at reconstructing 2D images from measurements that sample the object of interest with a series of spatially encoded patterns (structured light) and associating these known masks with the corresponding 1D intensity measured with a single detector. Such arrangement allows the production of low-cost imaging systems comprised of detectors with fine temporal resolution for lifetime calculations (for example PMTs, SPADs), hyperspectral adaptability and high sensitivity, thus permitting the concurrent generation of the desired high-dimensional datasets. Recently, single-pixel imaging's potential to simultaneously acquire spatial, temporal and spectral datasets at the macroscopic level has been demonstrated via the implementation of single-pixel based Macroscopic Fluorescence Lifetime Imaging (SP-MFLI). It integrates an ultra-fast laser, spatial light modulators, and a Time Correlated Single Photon Counting (TCSPC) spectrophotometer as described in [17]. This imaging platform enables the acquisition of dense time-domain data for functional imaging or lifetime-based molecular imaging for *in vivo* applications, both in 2D or 3D. [18–20] For a 2D formulation, which is used throughout this work, single pixel imaging retrieves the hyperspectral fluorescence time domain (TD) information by inverse solving Eq. (1) for the sample plane x at wavelength λ and time point t with m $(m \leq n)$ number of patterns where n represents the total number of pixels for a given resolution, $H_{m \times n}$ represents the pattern matrix and $b_{m \times 1}^{\lambda,t}$ represents the measurement vector. [17,21]

$$H_{m \times n}\, x_{n \times 1}^{\lambda,t} = b_{m \times 1}^{\lambda,t} \tag{1}$$

The m $(m \leq n)$ implies that the reconstruction of the sample plane x can take place with less patterns m and measurements b than the original required n amount of 1D measurements which equals the number of pixels at the desired resolution. [21]In order to comply to the m $(m \leq n)$ condition and alleviate the total acquisition time by reducing the number of acquisition patterns, compressive sensing (CS) strategies, based on sparse illumination basis and inverse solvers like the ubiquitously employed TVAL3 [22,23] are needed. Therefore, if only m patterned measurements out of n are used for reconstruction of the sample plane $x_{n \times 1}^{\lambda,t}$, then the data has been compressed by a compression ratio (CR) of:

$$CR = \frac{(n-m)}{n} \text{x } 100 \tag{2}$$

This means only *CR* percentage of patterned measurements are used for retrieving the image of the sample plane $x_{n \times 1}^{\lambda,t}$. Despite the multiple advantages of single-pixel arrangements like SP-MFLI and the option to reduce the amount of needed measurements $b_{m \times 1}^{\lambda,t}$ (and therefore acquisition time) through CS, one major drawback is the low reconstruction quality and accuracy of intensity and lifetime images at very high compression ratios (CRs), which to our knowledge have not been higher than 50% for resolution $n = 32 \times 32$ or 64×64 pixels in SP-MFLI reconstructions. [17,24] Such a requirement potentially leads to long acquisition times or ultra-weak signal conditions as well as long computational times and low resolutions. Therefore, a high compression ratio is desired since it would allow a decrease in acquisition time, making the SP-MFLI platform more suitable for fast *in vivo* imaging, while maintaining or enhancing pixel resolution. Moreover, due to the ill-posed nature of the inverse solvers, it is typical to employ advanced regularization techniques, which can be dependent on *post hoc* selection of tuning parameters. In this regard, we have recently proposed the use of deep learning to simultaneously reconstruct intensity and lifetime single pixel images at high processing speed and without the use of parameter constrained inverse solvers. [24] However, this approach is still dependent on inputting at least 50% of the spatially coded measurements and therefore limited to low resolutions or high acquisition times. To overcome this obstacle and accelerate bed-side translation of SP-MFLI, we propose NetFLICS-CR (Network for Fluorescence Lifetime Imaging with Compressive Sensing at high Compression Ratio), a deep learning framework that aims to reduce the number of required

1D measurements and acquisition times for SP-MFLI by more than one order of magnitude while increasing resolution to 128×128 pixels. Herein, NetFLICS-CR is tested and validated within the framework of preclinical small animal studies. Since SP-MFLI can acquire dense spatial, spectral and temporal data cubes for multiplexed fluorescence molecular imaging, [17] enhancing the technique is intended to advance the drug development pipelines by enabling fast and simultaneous monitoring of multiple biomarkers at 128×128 resolution, as well as the quantification of targeted drug cellular delivery via the measurement of near infrared (NIR) labeled receptor-ligand engagement via Föster Resonance Energy Transfer (FRET). [25] First, we report on the design and *in silico* training as well as validation of NetFLICS-CR. Then, we establish NetFLICS-CR robustness by reconstructing controlled *in vitro* experimental data never used during the *in silico* training phase. Subsequently, NetFLICS-CR is used to quantify iron-bound transferrin (Tf) based target engagement in the liver and bladder of live, intact mice. Lastly, we report its applicability to monitor *in vivo* drug delivery via SP-MFLI FRET in mice bearing breast cancer tumor xenografts. Specifically, intracellular delivery of trastuzumab (TZM) was quantified in human epidermal growth factor receptor 2 (HER2)-positive tumors in live intact mice. TZM is an antibody that binds to HER2 on the surface of cancer cells. Notably, TZM is a FDA-approved drug for the treatment of metastatic breast and gastric cancer that overexpress HER2 [26].

2. Methods & results

2.1. SP-MFLI setup

The used SP-MFLI setup has been previously described in [17,21]. The system is used in its reflectance configuration with widefield illumination and structured light detection in the NIR range. The illumination is accomplished through a DMD-based pico-projector (PK101, Optoma, CA) by using a Mai Tai HP laser (Spectra Physics, CA) with 690 nm to 1040 nm pulsed excitation at 80 MHz repetition rate. The emissions from the sample plane are detected in a structured arrangement by a D4110 DMD (Digital Light Innovations, TX) and directed towards a fiber bundle connected to a Czerny-Turner type spectrophotometer (MS125, f-number: 3.7, Newport Optics, CA) that focuses the first-order diffracted light into a multi-anode PMT detector with active area of 16×16mm (PML-16-C, Becker & Hickl GmbH, Germany), consisting of 16 detection channels. The signals from the PMT are relayed to a TCSPC unit (SPC-150 and DCC-100, Becker & Hickl GmbH, Germany) to record the Time Point Spread Functions (TPSFs) generated from each of the detection channels.

2.2. TVRecon: TVAL-CS based reconstruction workflow

In order to evaluate NetFLICS-CR results versus a conventional CS based inverse solving method, the TVRecon (TVAL based reconstruction) approach is employed in parallel. TVRecon is a method for intensity and lifetime image reconstruction composed of two main algorithms, TVAL3 inverse solver and Least-Squares Minimization (LSQR) algorithm. TVAL3 stands for "Total Variation Minimization by Augmented Langranian and Alternating Direction Algorithms", which is a Matlab inverse solver that helps retrieve the image sample plane $x_{n x 1}^{\lambda, t}$ (Eq. (1)) from its degraded acquisitions. More information about TVAL3 mathematical model can be found in [22]. It is applicable to single-pixel imaging, where the sample plane is described by pattern matrix $H_{m x n}$ with sparsity constraints. SP-MFLI output data $b_{m x 1}^{\lambda, t}$ is composed of a set of Time-Point-Spread-Functions (TPSFs) and their recorded photon counts for each detection wavelength λ over 256 time points (t). One of these sets is generated per recorded pattern number m. The CS pattern basis used throughout this work will be the Hadamard Ranked basis, as it has previously demonstrated high performance for parallel acquisition of SP-MFLI data. [23] One important aspect of this workflow is that TVRecon is highly dependent on user defined parameters

for TD reconstruction of $x_{nx1}^{\lambda,t}$ through TVAL3. Furthermore, in order to obtain lifetime, each pixel TPSF in $x_{nx1}^{\lambda,t}$ must be subsequently fitted by a separate Least-Squares-minimization (LSM) algorithm for lifetime reconstruction. More information on the optimization for TD reconstruction through TVAL3 and further quantification of lifetime is provided in Appendix A.

2.3. NetFLICS-CR: architecture design

We recently demonstrated that SP-MFLI sequential image reconstruction tasks could be performed efficiently and simultaneously using deep learning (DL). Indeed, with NetFLICS [24], a Convolutional Neural Network (CNN), single pixel raw fluorescence data can be reconstructed into intensity and lifetime images in a single workflow and without the need of parameter optimization. Compared to the TVRecon workflow and when using 50% of the patterns for a 32×32 resolution, NetFLICS produced higher quality reconstructions even at low photon counts and at computational speeds 4 orders of magnitude faster. Despite these promising initial results, NetFLICS only reconstructs 32×32 images at a fixed compression ratio of 50% (512 measurements out of 1024). Even-though the increase in processing speed is essential to extend HMFLI for clinical *in vivo* settings, it is necessary to simultaneously increase resolution beyond 32×32 pixels while decreasing the acquisition time.

Herein, a new CNN architecture, NetFLICS-CR, as displayed in Fig. 1(a), is proposed to retain the fast processing speed for simultaneous intensity and lifetime retrieval but reconstruct 128×128 pixels while using compression ratios (CRs) of up to 99% to reduce acquisition time. For instance, for a 128×128 resolution, a 90% CR uses 1640 patterned measurements out of 16,384 while a 99% CR uses only 163 out of 16,384 or 1% m out of n measurements. (Eq. (1)) NetFLICS-CR is structured in a similar arrangement to NetFLICS [24], therefore retaining the three branched structure, with a common segment that derives into separate intensity and lifetime branches. Despite the structure similarity, the kernel sizes have been modified to account for the increase to 128×128 pixels, while for network training the common branch allows to input SP-MFLI simulated training data with different CR percentages. Moreover, to increase the training robustness on the lifetime branch, 2D Separable Convolutions [27] have been used to replace 2D Convolutions in order to better extract the lifetime features along the time dimension of the data. The intensity branch is composed of a single ResBlock [28] and ReconBlock [29] that yields the reconstructed 128×128 intensity image. The lifetime branch contains a ResBlock, a double ReconBlock structure and an additional 1D convolutional layer continued by a 2D separable convolution. The 1D convolutional layers are followed by batch normalization [30] and ReLU [31] activations. For training, the EMNIST dataset [32,33] is used as the spatial model to generate 128×128 pixels images. Then, at each pixel, fluorescence time-domain data is simulated using an exponential fluorescent decay model convolved with an HMFLI instrumental response function (IRF). The architecture and training process are detailed in Appendix B. For network training and validation per CR, 40,000 *in silico* TD samples were generated with 32,000 used in training (in batch sizes of 10 limited by GPU capability) and 8,000 in validation (a sample refers to an EMNIST image with an intensity and fluorescence decay range per pixel; therefore, a $128 \times 128 \times 256$ array). The single pixel measurements are generated by using Eq. (1). To establish the compression of NetFLICS-CR, only a subset m of the pattern basis H_{mxn} is employed for generating the single pixel measurements $b_{mx1}^{\lambda,t}$. In all cases, the Hadamard Ranked basis (patterns ranked from low to high spatial frequency) is used. [23] Therefore, a 99% CR corresponds to only using the first 1% ($m = 163$) of the ranked Hadamard basis, while 90% CR to using the first 10% ($m = 1640$).

The number of raw single pixel measurements considered per CR is varying as displayed in Appendix B. Note that the CR operates on the spatial dimension only. The training and validation for both intensity and lifetime branches as evaluated through Mean Absolute Error (MAE) are displayed in Fig. 1(b) per CR. Additionally, the process was repeated 5 times per CR to evaluate
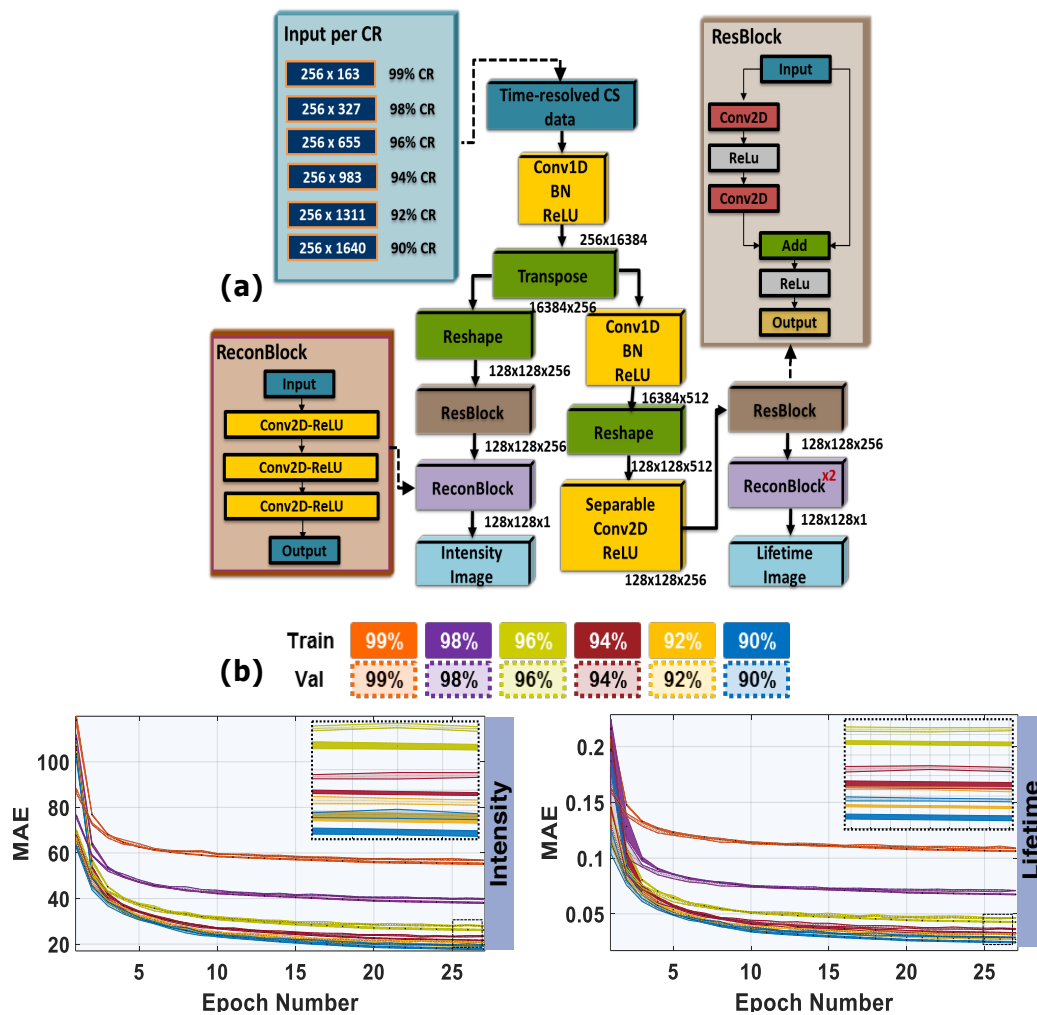
Fig. 1. a) Overall NetFLICS-CR block architecture: b) Mean Absolute Error (MAE) for intensity and lifetime training - validation curves at different CRs for 5 different training routines. Solid colors represent training curves while transparent ones represent validation curves. Epoch number was controlled through early-stopping during network training.

training stability. The results indicate, for both intensity and lifetime, that as the CR increases the MAE also increases, which is expected as there is less data to reconstruct from. The training and validation curves converge in similar MAE values for the 5 different trainings for each CR, as shown through the solid color shading for training curves and transparent color for validation curves (Fig. 1 (b)). The data generation and network training codes are available upon request to the authors.

2.4.   *In silico reconstructions and quantification*

To evaluate NetFLICS-CR accuracy in terms of image reconstruction at each compression ratio, 400 new TD *in silico* samples (different from the training and validation set), were reconstructed. The same samples were also reconstructed with the two-step TVRecon procedure, yielding an intensity and lifetime image per CR. The results of both methods were thresholded for background to 5% of the maximum intensity value and compared versus ground-truth through the Structural Similarity Index Matrix (SSIM) metric, which for an ideal reconstruction should equal 1. [34] The 128×128 intensity and lifetime reconstructions for the three highest compression ratios are displayed for each method in Fig. 2(a) for an example simulated data set. The average SSIM for the sample sets is shown in Fig. 2(b) for intensity and lifetime, NetFLICS-CR and TVRecon methods.
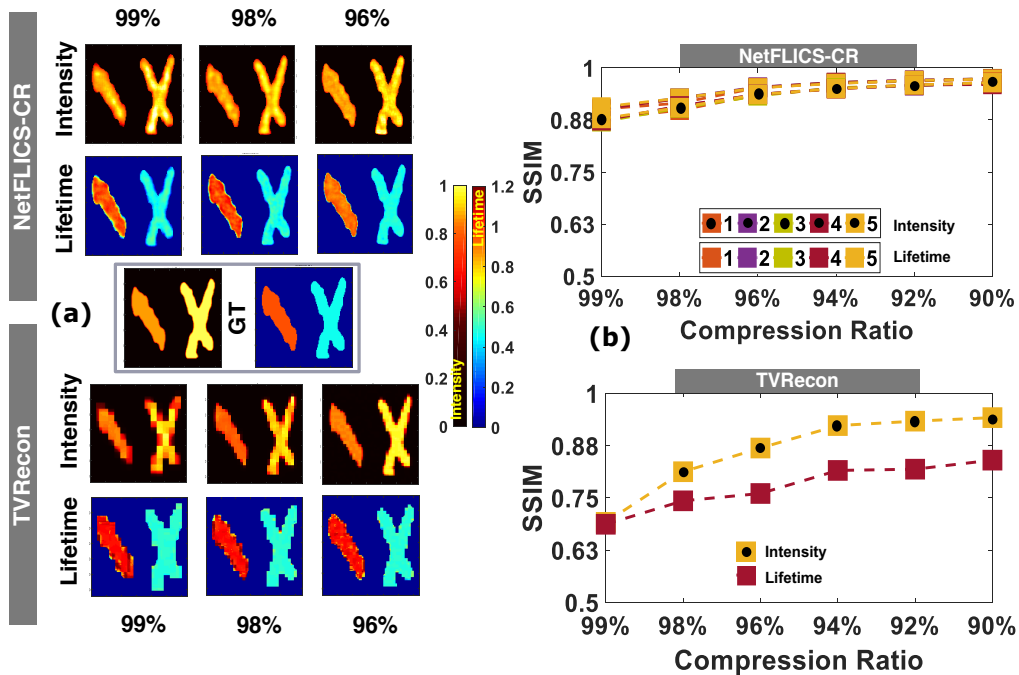
Fig. 2. a) Intensity and lifetime reconstructions versus ground-truth (GT). One of 400 reconstructed samples displayed for visualization. b) Average intensity and lifetime SSIM values calculated versus ground-truth at different CRs: 90%-1,640, 92%-1,311, 94%-983, 96%-655, 98%-327 and 99%-163 out of 100%-16,384 patterns). For NetFLICS-CR the legends 1-5 indicate results for the same *in silico* sample-set for five different trainings. Intensity and lifetime colorbars in (a) apply all reconstructions.

Since NetFLICS-CR was trained 5 times per compression, one SSIM value per each training and CR is plotted. SSIM values for both intensity and lifetime are higher than 0.8 at 99% CR and the SSIM values increase close to 0.9 as the CR decreases to 90%, which is expected due to an increase in acquired signal. Of note, SSIM values for lifetime are higher than those for intensity. This might be attributed to the higher range in simulated intensity values compared to lifetimes (also representative to experimental conditions). On the other hand, for all cases, TVRecon SSIM values are below 0.8 at all CRs. Note that the TVRecon method requires selection of the best regularization parameters, which were optimized to values yielding the highest SSIM intensity reconstructions versus ground-truth. After optimization and TD reconstructions, LSM was applied to yield lifetime images. This process is shown and further explained in Appendix A. Overall, the *in-silico* results with known ground-truth reconstructions indicate a better performance for NetFLICS-CR for both intensity and lifetime reconstructions with SSIM values always above 0.8.

### 2.5. *In vitro fluorescence reconstructions and quantification*

To test the experimental performance of NetFLICS-CR *in vitro* and the effect of compression on overall acquisition time, a phantom composed of continuous letters "R", "P" and "I" was used. Details of the phantom preparation are provided in Appendix C. Letters R and I contained Alexa Fluor 750 (AF750; ThermoFisher Scientific, A33085) dye while letter P was filled with HITCI (Sigma Aldrich, 252034). The expected lifetime values have been previously reported according to studies with these dyes and buffers. Expected average values of 0.50 ± 0.017ns for AF750 and 0.92 ± 0.018ns for HITCI are described in [23] for lifetime quantified with an external gated-ICCD camera. Additionally, values of 0.48 ± 0.02ns for letter R, 0.47 ± 0.03ns for I with AF750 and 0.84 ± 0.07ns for P with HITCI are reported in [24]. Therefore, on average, expected values are 0.49 ± 0.022 for RI-AF750 and 0.88 ± 0.044 for P-HITCI. The phantom was excited at 740 nm with 24 mW/cm^2 power on a field of view of 35×35mm.The samples were acquired with the SP-MFLI system using an acquisition time of 0.5 s per pattern, for a total of ~3 minutes for a 99% CR, ~6 minutes for a 98% CR and ~11 minutes for a 96% CR. Lower CRs were not considered as the main goal was to reduce experimental acquisition time through high data compression. Acquisition patterns were displayed on the digital micromirror device (DMD) as a positive and a 'negative' part. The negative part is the same pattern but with 0's inverted to 1's and vice-versa. Since Hadamard patterns are composed of 1's and -1's and the DMD is unable to display values below 0, the negative part will be used to calculate the correct signal as measured with the original patterns. Therefore, respective pattern numbers of 163×2 for 99% CR, 327×2 for a 98% CR and 655×2 for a 96% CR were acquired. The raw data was reconstructed using both TVRecon and NetFLICS-CR workflows. For TVRecon, Appendix A shows the primary and secondary penalty parameters that were used. For the lifetime quantification, TD pixels obtained from TVRecon with more than 5% counts of the maximum intensity value were fitted in order to obtain the lifetime values. Since this is a challenging minimization problem, the initial lifetime value was set to 0.8 ns and the fitting range bounded to [0.2-1.4] ns. This range covers the expected lifetime ranges of both letters at the same time. For NetFLICS-CR the *in silico* trained network was used at each CR to reconstruct both intensity and lifetime images (i.e. the network was not trained using experimental data). The reconstruction time of NetFLICS-CR was ~10 s for 16 spectral channels, while the full TVRecon approach took ~80 s per spectral channel (for a total of ~21 minutes). After intensity reconstruction, for both methods, pixels with less than 20% of the maximum intensity value were set to 0 as background threshold. Images were then normalized to the maximum value for SSIM comparison to an experimental intensity ground-truth (NIR CCD image of the sample plane).

The results for intensity and lifetime reconstructions for both methods are displayed in Fig. 3(a). As shown in Fig. 3(b), SSIM values at 99%, 98% and 96% CRs are higher for NetFLICS-CR

than for TVRecon reconstructions. In contrast to intensity, no lifetime ground-truth can be obtained in experimental settings. However, it is expected based on the experimental design that the lifetime will be homogeneous within each letter. Additionally, letters R and I should yield a similar lifetime as they both contain AF750 dye from the same stock solution. Results for lifetime reconstructions are displayed as histograms with a distribution per CR and method in Fig. 3(c). Two marked histogram peaks are located close to average "expected" lifetimes of 0.49 ± 0.022 for AF750 (letters R and I) and 0.88 ± 0.044 for HITCI (letter P). In contrast to TVRecon, at these CRs, NetFLICS-CR resulted in sharper distributions around the average "expected" values indicating a more accurate lifetime quantification per fluorophore. Note that the pixels used for reporting the reconstructed lifetimes are the ones used for the intensity comparison. Hence, these *in vitro* results indicate that NetFLICS-CR, even if only model trained, retrieves more accurate lifetime and intensity values when applied to experimental data sets at high CRs and 128×128 resolution. Additionally, NetFLICS-CR leads to a significant reduction in the acquisition time.
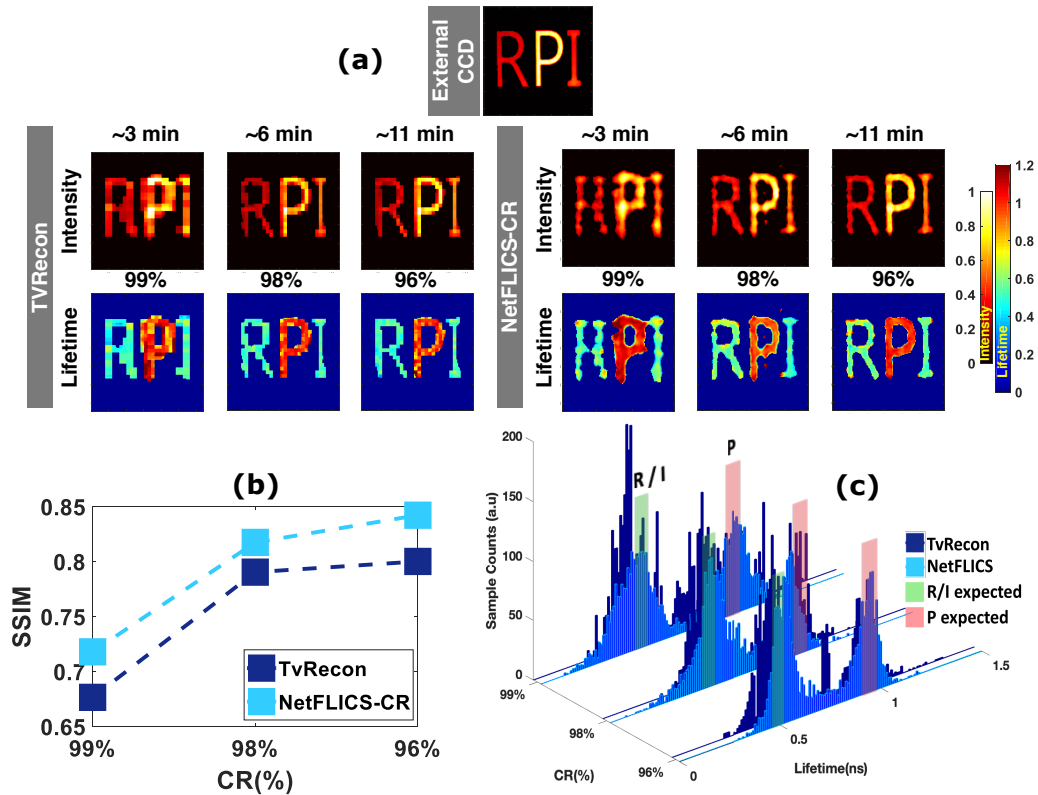
Fig. 3. a) *In vitro* SP-MFLI Experiments for RPI phantom: a) Intensity and lifetime reconstructions for RPI phantom per CR and acquisition time. Reconstructions for NetFLICS-CR and TVRecon methods displayed b) SSIM values calculated comparing reconstructed intensity images per method and CR versus external CCD 'ground-truth'. c) Distribution of lifetime reconstructions per CR and method for RPI phantom. Two main lifetime distributions observed (P and R/I). Expected values based on [23,24] for same dyes and buffer. Intensity and lifetime colorbars in (a) apply for both TVRecon and NetFLICS-CR reconstructions.
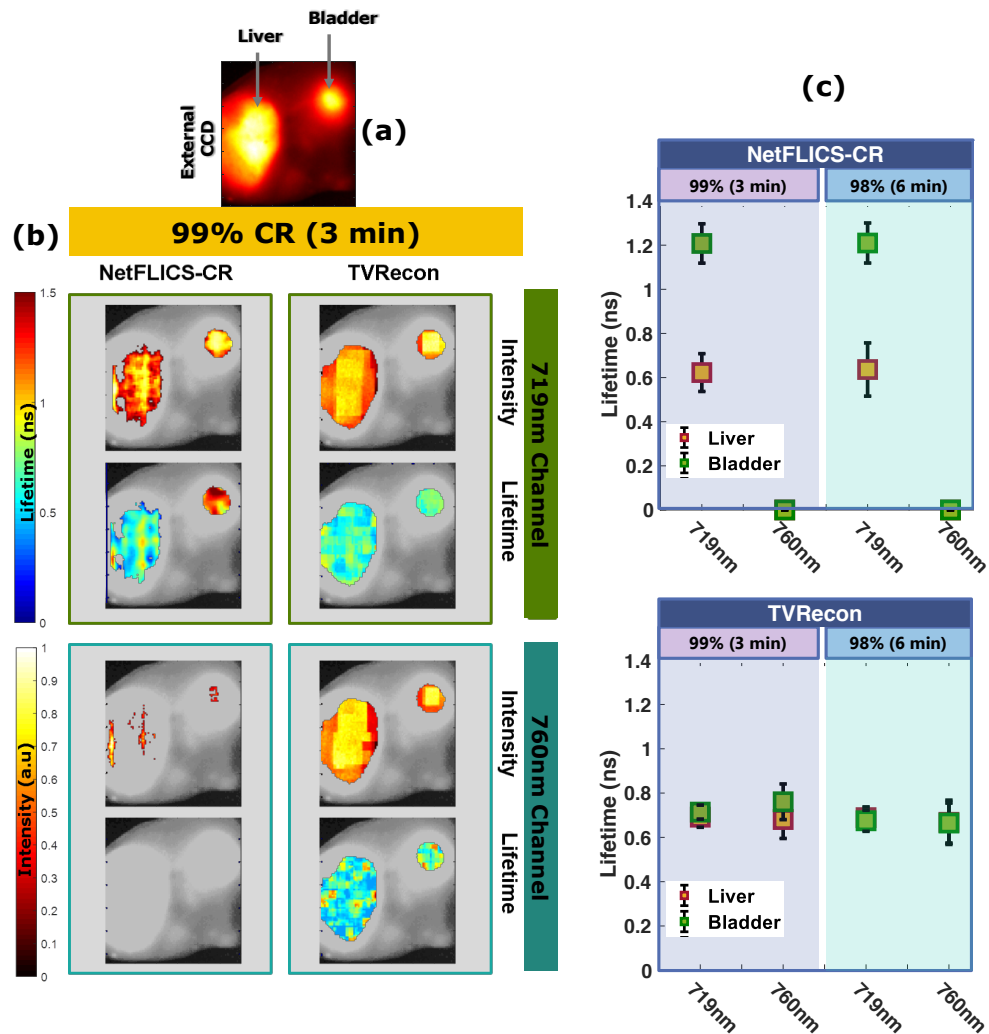
Fig. 4. a) Liver and bladder areas for the external CCD image of the sample plane. b) Intensity and Lifetime reconstructions for Tf-AF700 Mouse experiments for 99% CR (3 minutes acquisition). Detection at ~760 nm is displayed as a control for reconstruction (little expected fluorescence) c) Average lifetime for liver and bladder per CR and reconstruction method is summarized at ~719 nm or ~760 nm channels. Intensity and lifetime colorbars apply for both 719 nm and 760 nm.

2.6. *In vivo Transferrin uptake at organ level*

For the first *in vivo* validation the performance of NetFLICS-CR was tested in reconstructing receptor-ligand engagement of a Transferrin (Tf) – AF700 conjugated probe injected in a live intact mouse. The probe is designed to bind to Tf receptors located in the liver. *In vivo* sample preparation for imaging is further explained in Appendix D. The mouse was imaged using the SP-MFLI platform after 6 hours post-injection at a minimum CR of 98% with ~6 minutes acquisition time at 700 nm and 31 mW/cm^2 excitation for a 38×38 mm FOV. Therefore, 99% CR reconstructions could be later retrieved to represent acquisitions of ~3 minutes, respectively. Results for the Tf-AF700 mouse experiments for a 99% CR are displayed in Fig. 4 and in Appendix E (Fig. 9) for 98% CRs. The liver is a major site for iron homeostasis, so its Transferrin (Tf) receptor levels are high, leading to significant Tf uptake [35]. In addition, due to the liver's detoxifying function, changes in its microenvironment, such as in pH and ion composition, consistently result in a significant quenching. In contrast, urinary bladder, an excretion organ, should not cause donor fluorescent lifetime decrease, as shown by previous studies. [35–37] Both TVRecon and NetFLICS-CR methods were employed for comparison. NetFLICS-CR intensity and mean lifetime images were directly outputted from the network at CRs of 99% and 98%.

The same trainings used for *in silico* and *in vitro* SP-MFLI experiments have been used for *in vivo* Tf-AF700 NetFLICS-CR reconstructions. On the other hand, TVRecon intensity images were reconstructed by TVAL3 inverse solver and mono-exponentially fitted for lifetime (as only Tf-AF700 probe is being used) through LSQR with initial values of 0.9 ± 0.5 ns. Reconstructions have been overlaid over a grayscale intensity image of the FOV acquired with an external CCD camera. In order to correctly quantify and compare between CRs and methods, the regions of interest (liver and bladder) have been set as constant using the external CCD "ground-truth" image in Fig. 4(a). For this image, the values below 50% of the maximum intensity are set to zero as a background threshold. Then it is turned into a binary image that will multiply each of the reconstructions per CR/reconstruction method. This way the number of pixels in the desired region of interest are constant. Therefore, a total of 483 pixels are selected for bladder lifetime quantification and 2251 pixels for the liver. Final intensity and lifetime images are displayed in Fig. 4(b). Each intensity image is normalized to its own maximum value. The average lifetime and standard deviation of liver/bladder is shown in Fig. 4(c) per CR and reconstruction method.

TVRecon results display little to none lifetime change between liver and bladder for both CRs at 719 nm and 760 nm wavelength channels. Though, previous studies have reported on the quenching of Tf-AF700 in the liver due to the high levels of TfR-Tf binding in hepatic cells. Conversely, the bladder is an intra- and inter-subject control as Tf-AF700 quenching is reduced due to excretion of degraded Tf and free dye. For TVRecon the quantified lifetime values fall below the expected lifetimes for Tf-AF700 [24,36,38,39]. In contrast, NetFLICS-CR which describes an average lifetime of 1.2 ns for the bladder (unquenched) and a decrease in Tf-AF700 lifetime for the liver (quenched), provide outputs that are in accordance with prior work. Moreover, at the 760 nm control channel, NetFLICS-CR shows little to none intensity and lifetime reconstructions, while TVRecon shows intensity reconstructions for the liver and bladder regions, as well as similar lifetime values to the 719 nm channel for both organs. From Appendix F (Fig. 10) the first widefield raw measurement across wavelength channels is displayed, with channel 3 representing 719 nm and 13 representing 760 nm. Here the photon counts recorded for 760 nm are 12 times less than the 716 nm peak channel for Tf-AF700. However, when these counts are integrated over time channels, which is what TVRecon does for obtaining the 2D intensity reconstruction, it might be possible that intensity is retrieved as the counts difference is reduced to 6.5 times below the max peak counts. Thus, even though NetFLICS-CR reconstructs some intensity pixels at the 760 nm channel, it would benefit from including a lower count training set with characterized system noise, which are the further investigative steps for this architecture. Since the TPSFs per pixel in TVRecon are inverse solved through TVAL3 and then fitted through

Least-Squares minimization, we have displayed in Appendix F how the TVAL retrieved TPSF looks like for the maximum intensity pixel located at [X = 98, Y = 106] for the time domain reconstruction at 760 nm. Even though a fit is obtained the TPSF is highly noisy compared to the IRF. Therefore, it means that either a wider range of parameters besides μ and ß need to be optimized to inverse solve the best signal to noise ratio TPSF or the LSQR minimization algorithm needs further optimization to discard fits with "unacceptable" residuals. In this regard, we believe NetFLICS-CR might be more accurate at describing the 760 nm channel.

2.7. *In vivo TZM drug uptake in tumors*

We finally investigated the performance of SP-MFLI + NetFLICS-CR in a very challenging scenario, the quantification of TZM binding to a HER2-positive [38] AU565 tumor xenograft in a live intact mouse, upon intravenous injection of TZM-AF700 (Donor) and TZM-AF750 (Acceptor). The mouse was imaged at 24 and 102 hours post-injection to obtain insights on whether the technique could localize tumors and quantify *in vivo* drug uptake in dim conditions. AF700-TZM and AF750-TZM FRET pair was intravenously injected at a 2:1 acceptor to donor ratio in an athymic nude mouse bearing an AU565 tumor xenograft. Further explanation of the xenograft preparation is shown in Appendix D. The mouse was imaged *in vivo* with a minimum CR of 98%. The excitation was set to 31 mW/cm^2 at 700 nm for a 38×38 mm FOV, yielding a total acquisition time of ~6 minutes (256 TD temporal bins and 16 spectral channels). Intensity and lifetime images are reconstructed per time point for both donor and acceptor peak channels at respective detection wavelengths of ~719 nm and ~760 nm. NetFLICS-CR intensity and mean lifetime reconstructions were directly resolved, while TVRecon TD intensity data was inverse-solved by TVAL3 as explained in Appendix A. For TVRecon lifetime, each pixel was bi-exponentially fitted to return $A1$, $A2, \tau_1$ and τ_2 values, which respectively represent the percentage of FRETing Donor (FD%), Acceptor and their short and long lifetime components. To directly compare to the mean lifetime output of NetFLICS-CR, mean lifetime τ_{mean} was calculated through Eq. (4). After reconstruction with both methods, intensity images were normalized and the region of interest defined by an external CCD intensity image of the sample plane. The resulting region of interest was also applied for lifetime reconstructions. FRET occurring between donor and acceptor fluorophores would lead to quenching of the donor intensity and reduction of donor lifetime by the acceptor at the targeted tumor area. [39] Therefore, accessing both donor and acceptor channels could provide insights on the level and distribution of FRET events within the tumor. NetFLICS-CR and TVRecon reconstructions at the tumor xenograft area per each time-point are shown in Fig. 5(a) for 99% CR. Quantification for the mean lifetime values at the tumor region per post-injection time, reconstruction method, detection channel and CR is summarized in Fig. 5(b) and full set of reconstructed images displayed in Appendix G (Fig. 11).

In order to control the descriptive statistics, means account for the highest 80-pixel values in the tumor area per method and reconstruction type. According to the *in vitro* hyperspectral behavior of Tf-AF700 and Tf-AF750 and previously reported levels of FRET interaction between them, it is expected that the donor mean lifetime (expected value of ~1 ns) will decrease as it is quenched by the acceptor (expected value of ~0.5 ns), but the acceptor lifetime would minimally change. [17,24,40] In order to validate the expected values for the peak wavelengths, individual AF700 donor and AF750 acceptor probes were quantified *in vitro* along the 16 detection channels as displayed in Appendix H (Fig. 12) where lower detection channels are dominated by donor emission, while upper ones by acceptor emissions. Concentrations of AF700 and AF750 followed the previously described 20 μg and 40 μg concentrations used to yield a 2:1 ratio. Both average intensities and lifetimes are displayed per each detection channel ranging from channel 1 at 715nm to channel 16th at 783nm, with 4.5nm wavelength space between channels.

As shown in Fig. 5(b), at 24 hours p.i. in the tumor region there is a decrease in lifetime from donor to acceptor channel. Considering ligand/target engagement in the xenograft region to be
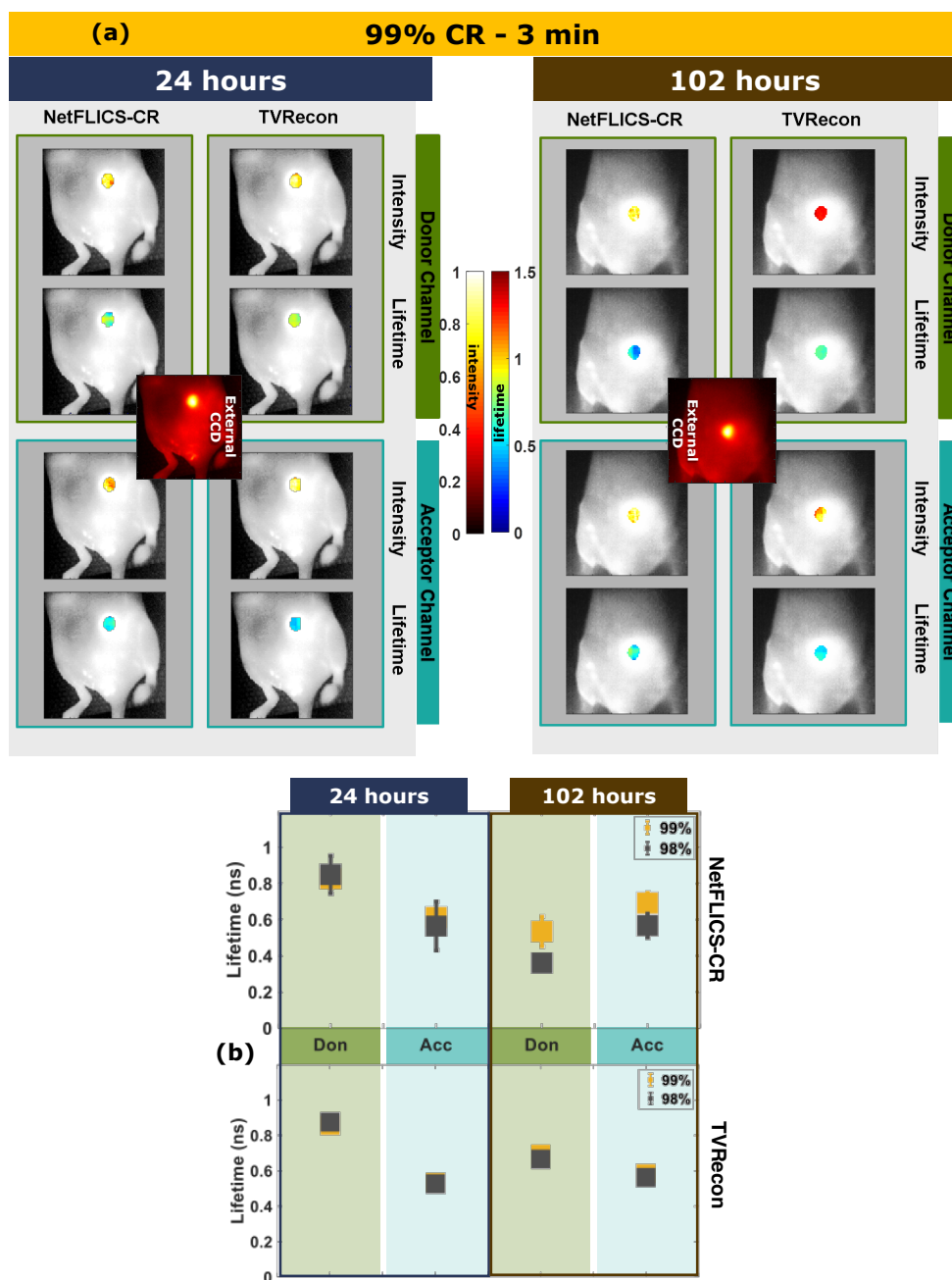
Fig. 5. *In Vivo* TZM drug engagement: a) Intensity and lifetime reconstructions for the tumor region at 24- and 102-Hours post-injection at 99% CR. b) Mean lifetime and standard deviation for Donor and Acceptor channels per time point, reconstruction method and CR. Lifetime and intensity colorbars in (a) apply for both 24 and 102 hours reconstructions.

heterogenous, NetFLICS-CR better approximates the expected values at 99% and 98% CRs, where TVRecon shows regions with no variation, which could indicate an over-regularization despite optimal parameter selection. At 102 hours p.i. donor quenching is expected due to the binding of donor and acceptor labeled TZM to HER2 dimerized receptors, resulting in a decrease in donor lifetime, as the drug undergoes internalization and endocytic trafficking. Conversely, since lifetime is independent of intensity and concentration, even if the signal of the acceptor decreases, lifetime should minimally change. Of note, from 24 hrs to 102 hrs both the donor and acceptor channel raw fluorescence signals decrease as displayed in Appendix I (Fig. 13). Since the Hadamard Ranked basis is organized by spatial frequency, Pattern 1 is expected to yield the signal with the most intense fluorescence decay. At 24 hours the TPSF at the donor channel is less intense than the TPSF at the acceptor channel. On the contrary, at 102 hours post injection, despite using the same acquisition settings used at 24 hours (excitation at 700 nm and 31 mW/cm^2 for a 38×38 mm FOV), both donor and acceptor TPSFs are below 200 photon counts. This is expected as the fluorescently tagged drug is excreted from the live animals leading to reduced local concentrations overtime. Additionally, the decrease in donor lifetime from 24 hrs to 102 hrs and acceptor lifetime (Fig. 5(b)) is suggestive of an increased fraction of AF700-TMZ undergoing FRET, i.e. intracellular delivery. This is expected since as the overall concentration of AF700-TMZ is decreasing due to excretion out of the animal, the intracellular fraction remaining is increasing over the extracellular fraction. Though, additional analysis and experiments of TZM engagement at the macroscopic level with immunohistochemistry validation are needed to have certainty of the amount of FRET expected at each timepoint. Last, even though Gaussian noise is included when simulating the fluorescent decays, we expect for future training sets to more accurately approximate the noise model and bi-exponential nature of the experimental TZM TPSFs; leading to overall improvements for lower photon count settings (tumor xenografts).

3. Discussion and conclusions

In conclusion, we report a novel CNN architecture, NetFLICS-CR, which efficiently reduces the acquisition time of high-dimensional SP-MFLI optical molecular data, while simultaneously producing 128×128 hyperspectral intensity and lifetime maps. Besides offering a fit free solution to the image formation paradigm, NetFLICS-CR led to a reduction in SP-MFLI acquisition times from ~2.5 hours at 50% CR to ~3 minutes at 99% CR. Despite the challenging photon-starved nature of *in vivo* acquisitions, NetFLICS-CR was able to reconstruct intensity and lifetime maps that were in accordance with the expected biological outcome. Additional benefits of NetFLICS-CR include it not requiring any user input optimization in contrast to fitting techniques and reconstructing intensity and lifetime images across 16 spectral channels in an approximate ~10s span compared to ~21 minutes needed for the TVRecon approach. This paves the way to a more standardized SP-MFLI platform for *in vivo* tissue characterization. It is worth noting that even-though a lifetime value can be directly calculated by fitting the raw TPSFs, to recover the intensity and lifetime spatial distribution it is necessary to retrieve it through the pattern weighted measurements either by a fit free paradigm as NetFLICS-CR or an inverse solver/fit based one like TVRecon.

Despite being able to describe the expected biological outcome, future studies with more mice are needed to understand the amount of uniformity expected from each lifetime region and how this varies across mice in both Transferrin and TZM based probes. Furthermore, it is necessary to further study the ideal intensity and lifetime-range for training NetFLICS-CR based on intensity and SNR variations of the targeted application, especially for very faint signals. Future work will pursue a further increase in resolution (beyond 128×128) as well as a decrease on the current ~3-minute minimum acquisition time (at 0.5s exposure per pattern) through optimizing the system's detection gating and binning parameters [41]. Beyond preclinical imaging, such a deep

learning paradigm is expected to greatly facilitate the translation of these new analytical tools to the clinical settings where acquisition and processing times are critical for patient intervention. Overall, this work illustrates how CNNs can play a central role in reducing the acquisition times for single-pixel multidimensional imaging at large, especially SP-MFLI. Additionally, we foresee that the NetFLICS-CR architecture herein described can be useful to guide other deep-learning developments in the field of compressed sensing and multiplexed imaging.
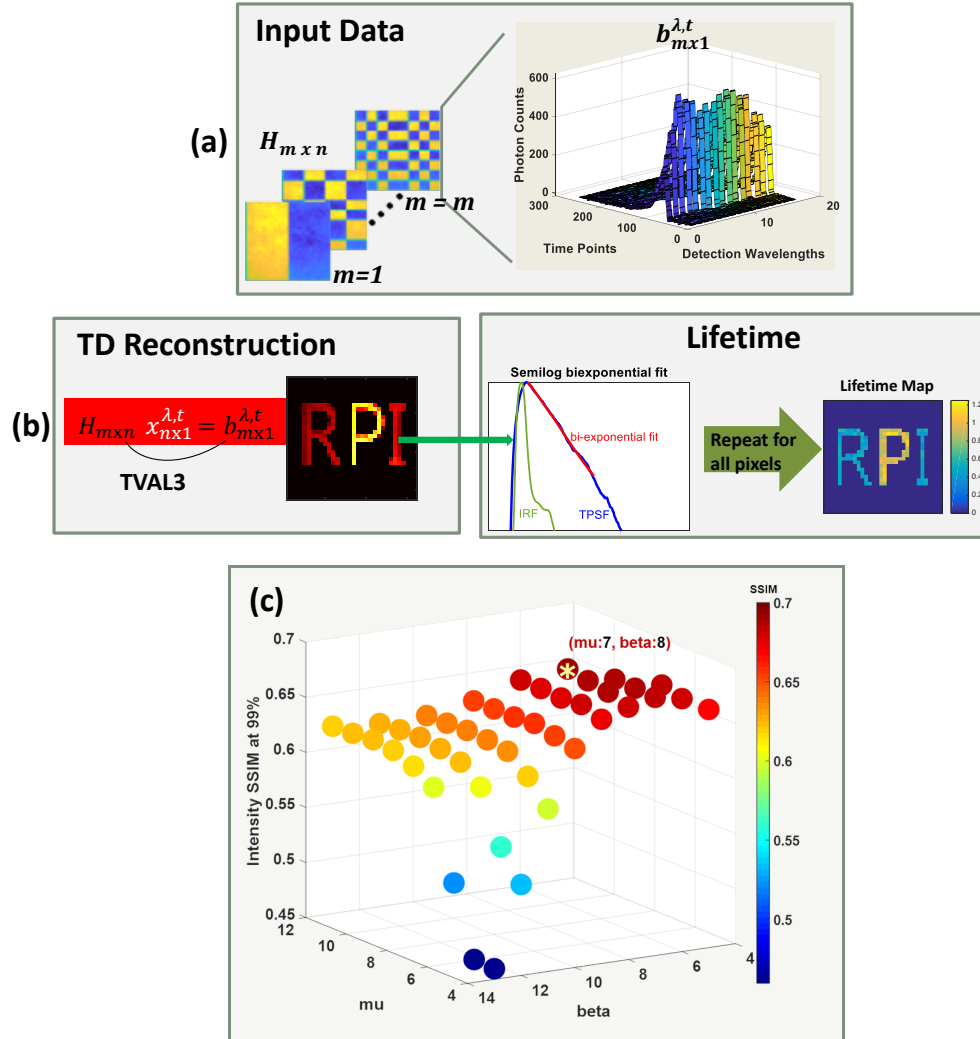
Fig. 6. (a) Input data for TVRecon and NetFLICS-CR methods composed of 1D measurements $b_{mx1}^{\lambda,t}$ per m number of patterns. Examples of Hadamard Ranked patterns displayed as H_{mxn} (b) Graphic description of TVRecon workflow with the first step inverse solving for $x_{nx1}^{\lambda,t}$ and the second the LSQR based fitting for each TPSF in $x_{nx1}^{\lambda,t}$ per n pixel (c) Optimization of primary μ(mu) and secondary β(beta) penalty parameters for TVAL3 based on intensity SSIM for 400 simulated samples.

Appendix A: TVRecon workflow

The data displayed in Fig. 6(a) is the structure of the input for both TVRecon and NetFLICS-CR. Further explanation of the optical system and its arrangement to yield this type of output is provided in [17]. For TVRecon, the input is translated into one TPSF per pixel of the desired resolution by inverse solving $H_{mxn}\, x^{\lambda,t}_{nx1} = b^{\lambda,t}_{mx1}$ for $x^{\lambda,t}_{nx1}$ as displayed in Fig. 6(b). The TVAL3 regularization terms used throughout this manuscript follow the isotropic model. Since the algorithm is highly dependent on these terms, the primary and secondary penalty parameters, being the most important according to the developers [22], are tuned. According to the developers suggestion the primary penalty μ and secondary penalty β parameters should be ideally set between $2\hat{}4$ and $2\hat{}13$. Therefore, for the 400 samples used during simulation experiments, Time-Domain reconstructions were made for all possible combinations of μ and β parameters within the suggested range. The Continuous-Wave intensity reconstructions were later evaluated versus ground-truth through SSIM and the resulting values averaged for the full sample set. The combination yielding the highest SSIM value (closer to 1) was used for Time-Domain reconstructions. The average sample set SSIMs for the possible combinations are displayed in Fig. 6(c). Therefore, values of $μ = 2\hat{}7$ and $β = 2\hat{}\,8$ were employed. Depending on the type of fluorescent probe used for experiments, for the second part of the TVRecon method, LSQR (Least-Squares Minimization) will be used to produce a mono-exponential or bi-exponential fit to each pixel TPSF of $x^{\lambda,t}_{nx1}$ to retrieve the lifetime values per pixel. For a bi-exponential model:

$$TPSF = IRF * [A_1 exp(-t/\tau_1) + A_2 \exp(-t/\tau_2)] \tag{3}$$

The TPSF on each pixel in $x^{\lambda,t}_{nx1}$ represents the convolution (*) of the IRF with a bi-exponential (or mono-exponential) model composed of τ_1 short lifetime and τ_2 long lifetime components, in respective fractions of *A1* and *A2*, such that *A1 + A2 = 1*. Therefore, the mean lifetime can be obtained through:

$$\tau_{mean} = \left[\left(\frac{A_1}{100}\right)\, \tau_1 + \left(\frac{A_2}{100}\right)\, \tau_2\right) \right] \tag{4}$$

The optimization is done using Matlab's "fmincon" [42] where upper and lower bounds have to be initialized. The initialization values correspond to the long and short lifetimes in nanoseconds, which are estimated factory values. Each value has a parameter bound with +/- units. These values are specified in the main text for each experimental set.

## Appendix B: NetFLICS-CR training and specifics

Since the physics behind single-pixel measurement generation is known, NetFLICS-CR is trained on simulated single-pixel acquisitions. EMNIST [32,33] figures consisting of digits and letters are re-scaled from 28×28 pixels to the desired resolution, in the case of NetFLICS-CR, of 128×128 pixels. Furthermore, data is augmented by randomly clustering many of the rescaled EMNIST figures and then resizing the space to 128×128 by down-sampling. The images are rotated, flipped, rescaled and organized randomly across the 128×128 space so that no repeating figures exist. To mimic the single-pixel data generation, 128×128 intensity images randomly varying from 200 to 800 photon counts and their corresponding lifetime images with random values from 0.3 to 1.2 nanoseconds were used. These values aim to cover typical ranges obtained in the experimental sets. Since the time gates used in the experimental procedure are kept at 256 with intervals of 32.6 ps, a fluorescence decay curve with Poisson noise can be simulated for each pixel matching the given intensity and corresponding lifetime values of the "sample space". TPSFs were "acquired" by convolving the decay with an experimentally acquired IRF. To simulate single-pixel acquisition, weights with value -1 or +1 given by the set of used "illumination Ranked Hadamard" patterns [23], are applied to the sum of all TPSFs from the "sample space". As using the full pattern basis for a 128×128 resolution is not experimentally viable, a total of

1800 Ranked Hadamard patterns were used for data generation in order to cover a minimum CR of 90% for training. An example of these sets is shown in Fig. 7(b). A total of 40,000 sets were used, 32,000 for training and 8,000 for validation. Each sample took approximately 2 hours to generate through MATLAB. After the samples are generated, the training per compression ratio is achieved by modifying the "pattern dimension" of the CW part of the data set as exemplified in Fig. 7(a) for different CRs. The common branch of NetFLICS-CR translates the input of size CNx256xPN, where PN represents the pattern number and CN number of detection channels, into 2D temporal data of dimensions 256×16384, which corresponds to the number of pixels in a TD 128×128 space.
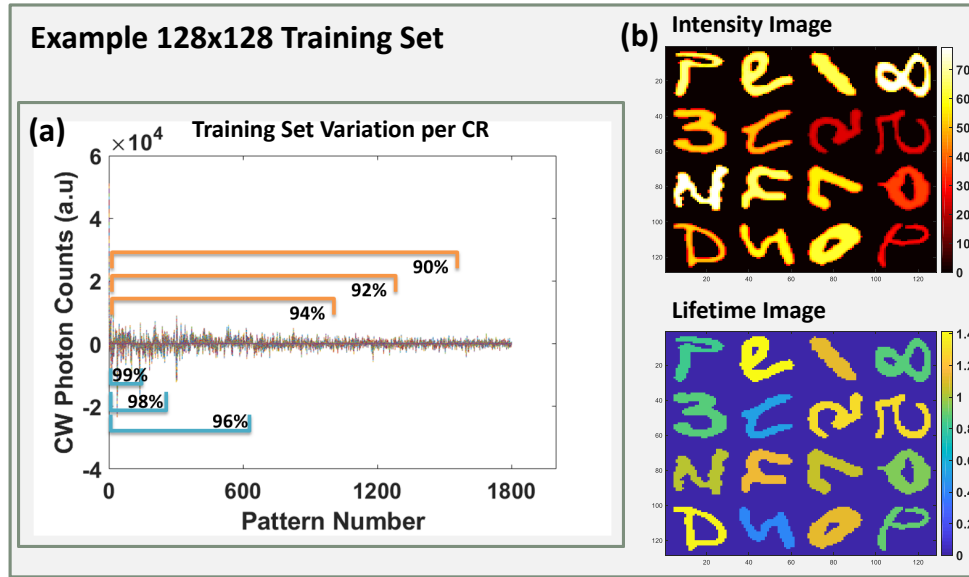
Fig. 7. (a) Continuous Wave visualization of simulated raw data and parts of the measurement that will be used to retrieve reconstructions for each compression ratio. (b) Example of simulated intensity and lifetime image corresponding to (a) for 128×128 resolution.

Additionally, in this branch, sparsity features are extracted from the input, through a 1D convolutional layer with 16384 size one 1D convolutional kernels that operate along the time dimension. Then batch normalization and ReLU activations are used. The output of segment 1 is permuted to 16384×256, so that it can be reshaped into the intensity branch to $128 \times 128 \times 256$, which is one TPSF per pixel. One ResBlock of 256 kernels with size 3×3 is followed by a ReconBlock formed by respective kernel numbers and sizes of 64 and 1×1, 32 and 1×1 and 1 and 3×3, which results in the 128×128 intensity image per detection channel. Parallel to intensity reconstruction, the transposed output from the common segment is received by a 1D convolution with 512 kernels of size 1 and batch normalization/ReLU activation. The output is further reshaped into $128 \times 128 \times 512$, which is the input for a separable 2D convolution with 256 kernels of size 1×1 and followed by a ReLU activation [43]. Then a ResBlock and two ReconBlocks are used to yield the 128×128 lifetime image per detection channel. Using an NVIDIA Titan XP GPU, NetFLICS-CR took an approximate of 17 hours for total training for 6 CRs (99%, 98%, 96%, 94%, 92%, and 90%).
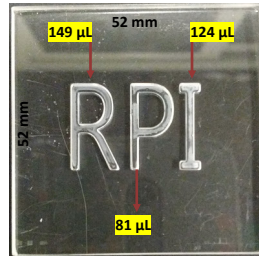
Fig. 8. RPI phantom dimensions and volume capacities per letter. For imaging it is centered within the 35×35 mm FOV of the SP-MFLI system.

Appendix C: *in vitro* experiments and sample preparation

The RPI phantom displayed in Fig. 8 is composed of continuous letters "R", "P" and "I" with respective volume capacities of 149, 124 and 81 μL. Alexa Fluor 750 (AF750; ThermoFisher Scientific, A33085) dye was prepared with PBS at 2.08 μM concentration to fill Letters "R" and "I", while HITCI (Sigma Aldrich, 252034) at 40 μM was used for letter "P". The initial concentration of HITCI was monitored versus AF750 intensity under an external NIR CCD camera and was diluted with ethanol until matching the intensity of both dyes. After filling the phantom with the fluorescent dyes, it was covered with a thin layer of clear wrap paper to minimize evaporation effects, which considering the intended acquisition times should be minimal.

## Appendix D: *in vivo* sample preparation and additional results

D.1 Ligand labeling. Human Holo Transferrin-Tf (Sigma Aldrich) and Trastuzumab-TZM (Genentech) were conjugated to AF700 donor only for the Holo Tf probe and both AF700 donor and AF750 acceptor (Life Technologies) for the TZM probe, through monoreactive N-hydroxysuccinimide ester to lysine residues in the presence of 100 mM Na bicarbonate, pH 8.3, according to manufacturer's instructions. The probes were purified by desalting columns and Amicon Ultra-4 microconcentrators (Millipore). The degree of labeling of the probes was assessed by spectrophotometer DU 640 (Beckman Coulter, Fullerton, CA, USA). The average degree of labeling was no more than 2 fluorophores per Tf or Trastuzumab molecule. All probes were normalized to concentration 1 mg/mL in phosphate-buffered saline pH 7.6 and filter sterilized.

D.2 Cell culture. Breast cancer cell line AU565 was purchased from ATCC (CRL-2351) and cultured in RPMI 1640 medium supplemented with 10% FBS in 5% CO_2 at 37°C in a humidified incubator for less than 12 passages.

D.3 Animal experiments. All animal procedures were conducted with the approval of the Institutional Animal Care and Use Committee at both Albany Medical College and Rensselaer Polytechnic Institute. Animal facilities of both institutions have been accredited by the American Association for Accreditation for Laboratory Animals Care International. Tumor xenografts were generated by injecting 10×10^6 AU565 cells in phosphate-buffered saline (PBS) mixed 1:1 with Cultrex BME (R&D Systems Inc, Minneapolis, MN, USA) into the right inguinal mammary fat pad of female 4-week old athymic nude mice (NU(NCr)-Foxn1[nu], Taconic Biosciences, Rensselaer, NY, USA). The subcutaneous tumors were allowed to grow for 4-5 weeks and were monitored daily. Tf-AF700 (40 μg in the volume 100 μL) or TZM-AF700 and AF750 (20 μg and 40 μg respectively) were injected retro-orbitally in anesthetized animals. For imaging, mice were anesthetized using E-Z anesthesia breathing machine and placed on a heating pad to maintain

body temperature. The depth of anesthesia was monitored by breathing rate and the absence of response to a foot pinch.

**Appendix E:** *in vivo* **TfAF700 additional results**
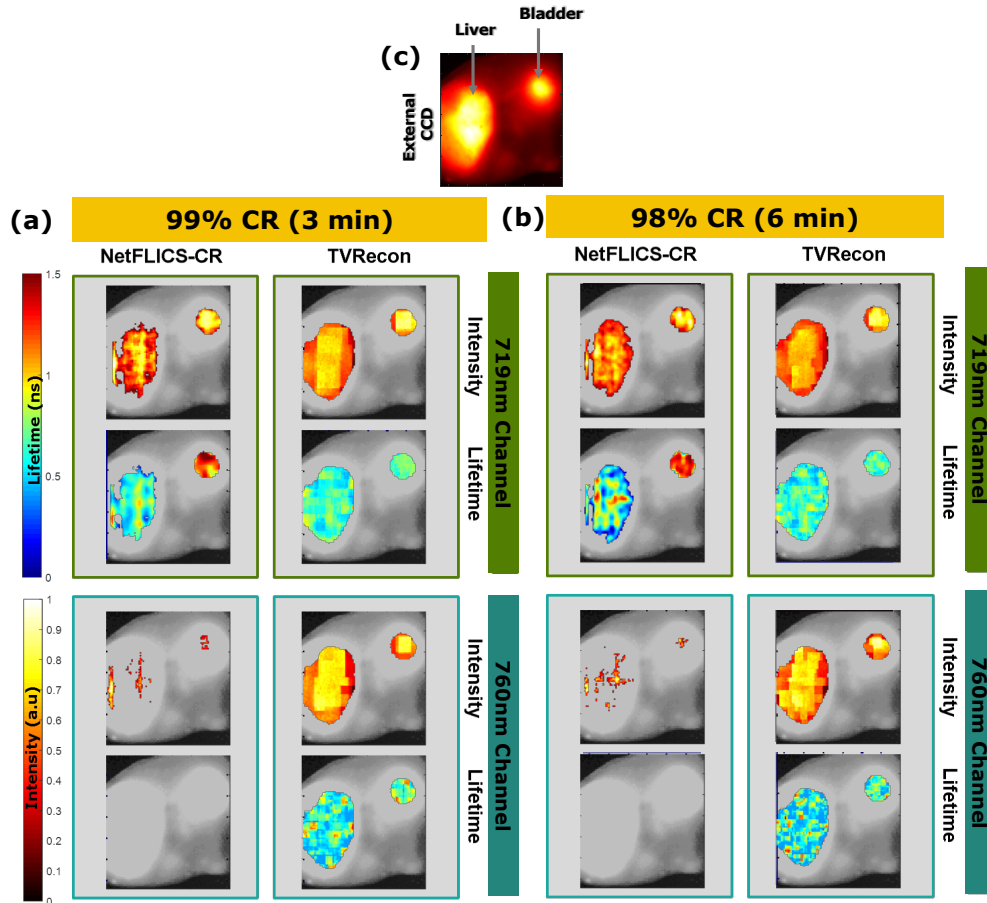
Fig. 9. Intensity and Lifetime reconstructions for Tf-AF700 Mouse experiments for 99% (a) and 98% (b) CRs for both NetFLICS-CR and TVRecon methods. Detection at the ~760 nm Channel is displayed as a control for reconstruction, since little detected fluorescence is expected at this wavelength. (d) Liver and bladder areas are respectively indicated in the external CCD image of the sample plane. Reconstructed images are overlaid over external CCD grayscale image of the mouse.

Appendix F: *in vivo* TfAF700 additional analysis of 760 nm control channel
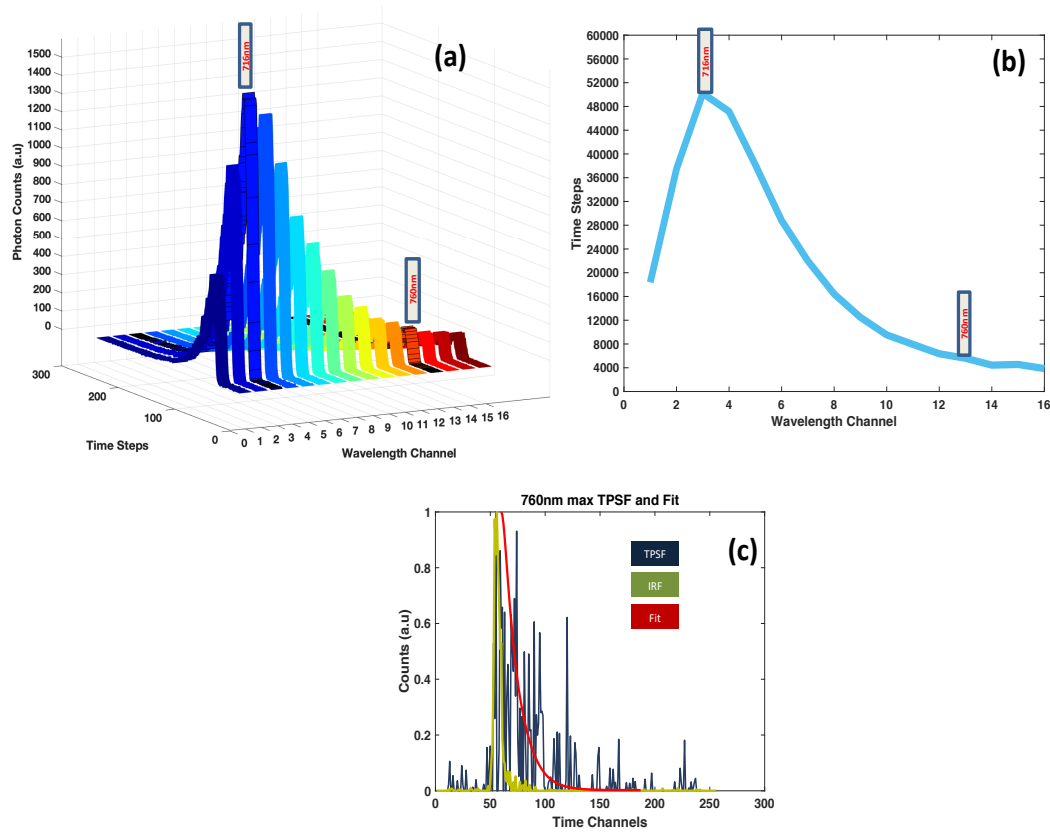
Fig. 10. (a) Pattern 1 widefield raw measurement across wavelength channels. Channel 3 represents 719 nm and channel 13 760 nm. (b) Continuous wave (over time) integration of measurement vectors per wavelength channel. (c) Retrieved TVAL3 TPSF for maximum intensity pixel (99%CR) located at [X=98,Y=106] for the time domain reconstruction at 760 nm. TPSF (blue), IRF (green) and LSQR fit (red) are displayed.

Appendix G: *in vivo* **TZM reconstructions for 99% and 98% CRs at the two different time-points.**
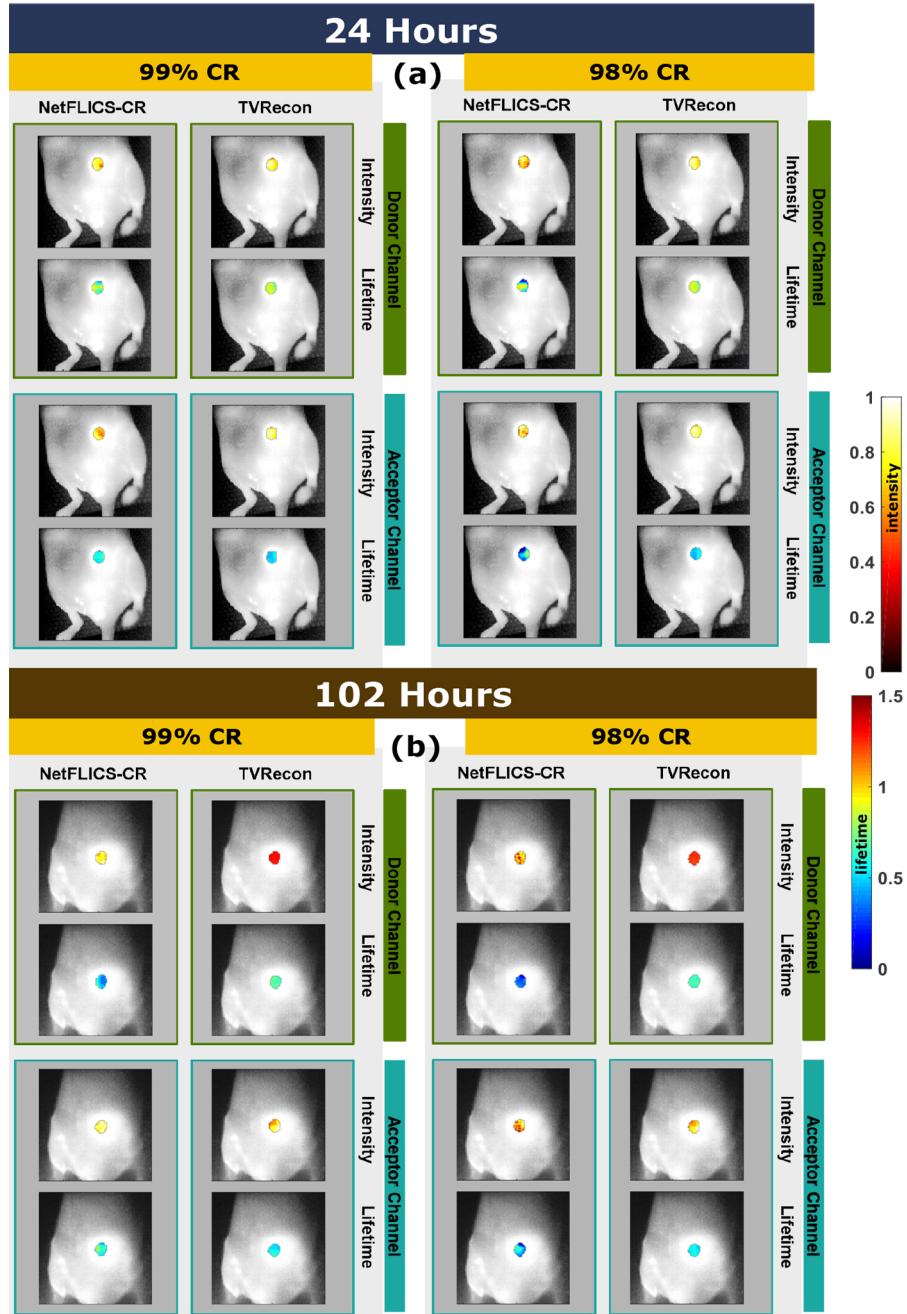

Fig. 11. Intensity and lifetime reconstructions for TZM injected athymic nude mouse with tumor xenograft at 24 (a) and 102 (b) hours post-injection. Reconstructions provided for both Donor (∼719 nm - green) and Acceptor (∼760 nm - Teal) detection channels for a 99% and 98% CR.

Appendix H: *in vitro* quantified AF700 donor and AF750 acceptor probes across detection channels.
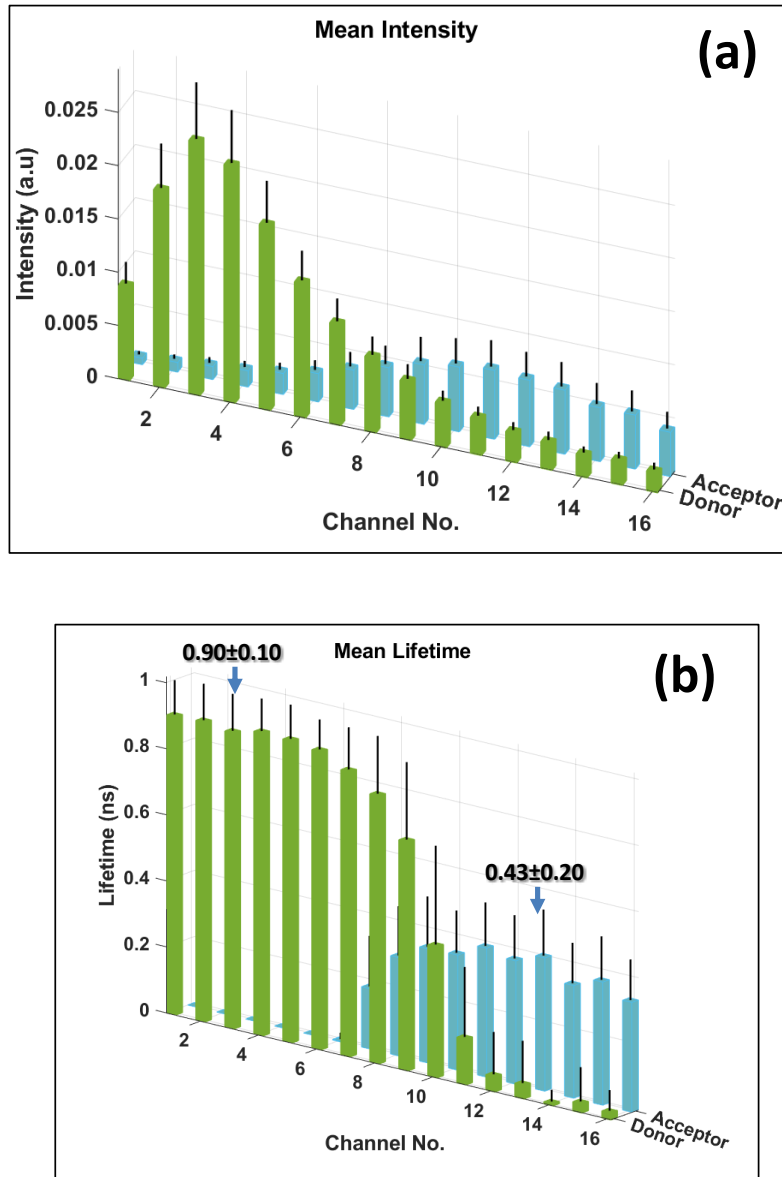
Fig. 12. (a) Average intensity values for both donor and acceptor. (b) Mean lifetimes and their standard deviations are shown per detection channel. Lifetime values in peak channels quantified for Donor (~719 nm at Channel 3) and Acceptor (~760 nm at Channel 13) are highlighted. Note even though AF700 quantification is within the same standard deviation from Channel 1 to Channel 8th at ~742 nm. After this wavelength little emission from AF700 is expected and therefore the lifetime quantification may be inaccurate. This also applies for Channels 1 to 6 for AF750 emissions.

Appendix I: TZM raw fluorescent signals across time-points.
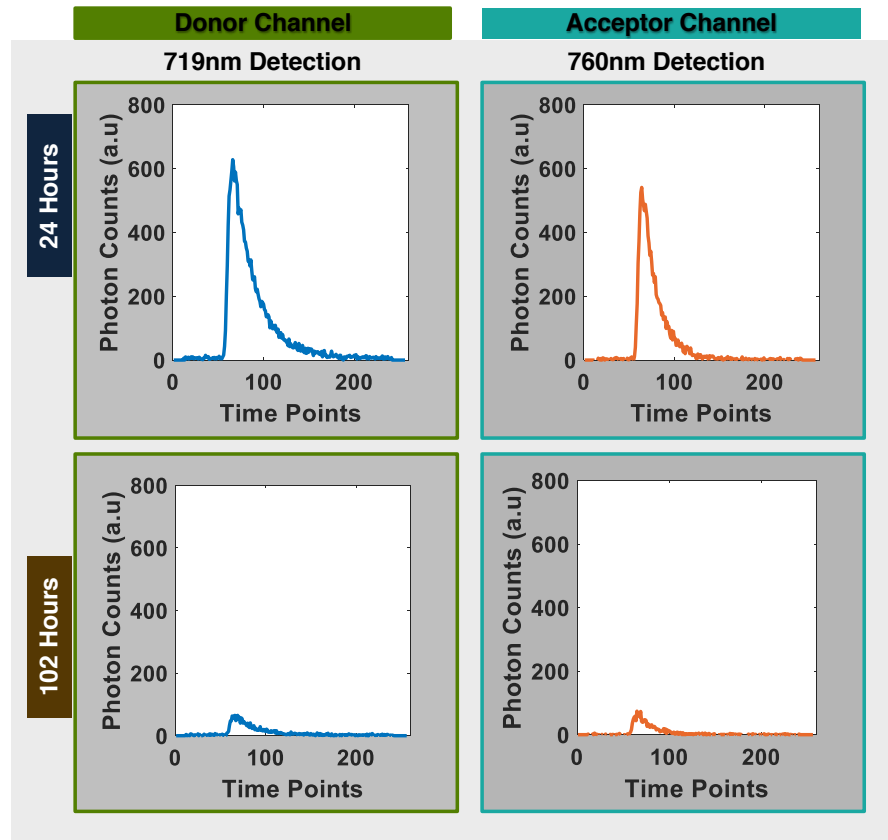
Fig. 13. Raw TPSF for Pattern 1 for both Donor and Acceptor channels at respective wavelengths of ~719 nm and ~760 nm. Since the Hadamard Ranked basis is organized by spatial frequency, Pattern 1 is expected to yield the signal with the highest fluorescence intensity. Pattern 1 TPSF is displayed per channel for 24 and 102 Hours raw TZM in vivo acquisitions. Note that this is raw data, therefore before NetFLICS-CR or the TVRecon inverse solver have recovered the spatial distribution of intensity or lifetime.

Disclosures

The authors declare no conflicts of interest.

References

1. J. N. Mait, G. W. Euliss, and R. A. Athale, "Computational imaging," Adv. Opt. Photonics **10**(2), 409 (2018).
2. X. Hu, J. Wu, J. Suo, and Q. Dai, "Emerging theories and technologies on computational imaging," Front. Inf. Technol. Electron. Eng. **18**(9), 1207–1221 (2017).
3. D. L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006).
4. F. Ströhl and C. F. Kaminski, "Frontiers in structured illumination microscopy," Optica **3**(6), 667 (2016).
5. J. P. Angelo, S.-J. K. Chen, M. Ochoa, U. Sunar, S. Gioux, and X. Intes, "Review of structured light in diffuse optical imaging," J. Biomed. Opt. **24**(7), 1 (2018).
6. M. G. L. Gustafsson, "Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution," Proc. Natl. Acad. Sci. **102**(37), 13081–13086 (2005).
7. W. Choi, C. Fang-Yen, K. Badizadegan, S. Oh, N. Lue, R. R. Dasari, and M. S. Feld, "Tomographic phase microscopy," Nat. Methods **4**(9), 717–719 (2007).
8. L. Waller, G. Situ, and J. W. Fleischer, "Phase-space measurement and coherence synthesis of optical beams," Nat. Photonics **6**(7), 474–479 (2012).
9. A. Orth, M. J. Tomaszewski, R. N. Ghosh, and E. Schonbrun, "Gigapixel multispectral microscopy," Optica **2**(7), 654 (2015).
10. W. Jahr, B. Schmid, C. Schmied, F. O. Fahrbach, and J. Huisken, "Hyperspectral light sheet microscopy," Nat. Commun. **6**(1), 7990 (2015).
11. S. Bélanger, M. Abran, X. Intes, C. Casanova, F. F. F. Lesage, S. Bélanger, M. Abran, X. Intes, C. Casanova, and F. F. F. Lesage, "Real-time diffuse optical tomography based on structured illumination," J. Biomed. Opt. **15**(1), 016006 (2010).
12. V. Venugopal, J. Chen, F. Lesage, and X. Intes, "Full-field time-resolved fluorescence tomography of small animals," Opt. Lett. **35**(19), 3189 (2010).
13. R. Strack, "Highly multiplexed imaging," Nat. Methods **13**(1), 35 (2016).
14. R. Ankri, A. Basu, A. C. Ulku, C. Bruschini, E. Charbon, S. Weiss, and X. Michalet, "Single-Photon, Time-Gated, Phasor-Based Fluorescence Lifetime Imaging through Highly Scattering Medium," ACS Photonics **7**(1), 68–79 (2020).
15. M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," IEEE Signal Process. Mag. **25**(2), 83–91 (2008).
16. M. P. Edgar, G. M. Gibson, and M. J. Padgett, "Principles and prospects for single-pixel imaging," Nat. Photonics **13**(1), 13–20 (2019).
17. Q. Pian, R. Yao, N. Sinsuebphon, and X. Intes, "Compressive hyperspectral time-resolved wide-field fluorescence lifetime imaging," Nat. Photonics **11**(7), 411–414 (2017).
18. Q. Pian, R. Yao, L. Zhao, and X. Intes, "Hyperspectral time-resolved wide-field fluorescence molecular tomography based on structured light and single-pixel detection," Opt. Lett. **40**(3), 431 (2015).
19. Q. Pian, R. Yao, and X. Intes, "Hyperspectral wide-field time domain single-pixel diffuse optical tomography platform," Biomed. Opt. Express **9**(12), 6258 (2018).
20. J. T. Smith, M. Ochoa, and X. Intes, "UNMIX-ME: spectral and lifetime fluorescence unmixing via deep learning," Biomed. Opt. Express **11**(7), 3857 (2020).
21. Q. Pian, "Time Resolved Hyperspectral Compressive Single-Pixel Wide-Field Optical Imaging," Rensselaer Polytechnic Institute (2017).
22. C. Li, "Compressive Sensing for 3D Data Processing Tasks: Applications, Models and Algorithms," RICE University (2011).
23. M. Ochoa, Q. Pian, R. Yao, N. Ducros, and X. Intes, "Assessing patterns for compressive fluorescence lifetime imaging," Opt. Lett. **43**(18), 4370 (2018).
24. R. Yao, M. Ochoa, P. Yan, and X. Intes, "Net-FLICS: fast quantitative wide-field fluorescence lifetime imaging with compressed sensing – a deep learning approach," Light: Sci. Appl. **8**(1), 26 (2019).
25. S. Rajoria, L. Zhao, X. Intes, and M. Barroso, "FLIM-FRET for Cancer Applications," Curr. Mol. Imaging **3**(2), 144–161 (2015).
26. J. Mendelsohn and J. Baselga, "The EGF receptor family as targets for cancer therapy," Oncogene **19**(56), 6550–6565 (2000).
27. F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* (2017).
28. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 770–778.
29. K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, "Reconnet: Non-iterative reconstruction of images from compressively sensed measurements," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 449–458.
30. S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *32nd Int. Conf. Mach. Learn. ICML 2015* 1, 448–456 (2015).
31. V. Nair and G. E. Hinton, "Rectified linear units improve Restricted Boltzmann machines," in *ICML 2010 - Proceedings, 27th International Conference on Machine Learning* (2010).

32. L. Deng, "The MNIST database of handwritten digit images for machine learning research [best of the web]," IEEE Signal Process. Mag. **29**(6), 141–142 (2012).

33. G. Cohen, S. Afshar, J. Tapson, and A. van Schaik, "EMNIST: Extending MNIST to handwritten letters," in *2017 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2017), pp. 2921–2926.

34. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Trans. on Image Process. **13**(4), 600–612 (2004).

35. S. Chen, N. Sinsuebphon, A. Rudkouskaya, M. Barroso, X. Intes, and X. Michalet, "In vitro and in vivo phasor analysis of stoichiometry and pharmacokinetics using short-lifetime near-infrared dyes and time-gated imaging," J. Biophotonics **12**(3), e201800185 (2019).

36. A. Rudkouskaya, N. Sinsuebphon, J. Ward, K. Tubbesing, X. Intes, and M. Barroso, "Quantitative imaging of receptor-ligand engagement in intact live animals," J. Control. Release **286**, 451–459 (2018).

37. N. Sinsuebphon, A. Rudkouskaya, M. Barroso, and X. Intes, "Comparison of illumination geometry for lifetime-based measurements in whole-body preclinical imaging," J. Biophotonics **11**(10), e201800037 (2018).

38. Y. Ardeshirpour, D. L. Sackett, J. R. Knutson, and A. H. Gandjbakhche, "Using in vivo fluorescence lifetime imaging to detect HER2-positive tumors," EJNMMI Res. **8**(1), 26 (2018).

39. K. Abe, L. Zhao, A. Periasamy, X. Intes, and M. Barroso, "Non-Invasive In Vivo Imaging of Near Infrared-labeled Transferrin in Breast Cancer Cells and Tumors Using Fluorescence Lifetime FRET," PLoS One **8**(11), e80269 (2013).

40. K. Abe, L. Zhao, X. Intes, and M. Barroso, "Quantitative Detection of Near Infrared-labeled Transferrin using FRET Fluorescence Lifetime Wide-Field Imaging in Breast Cancer Cells In Vitro and In Vivo," in *Imaging and Applied Optics* (OSA, 2013), p. QTu3G.5.

41. S.-J. Chen, N. Sinsuebphon, and X. Intes, "Assessment of Gate Width Size on Lifetime-Based Förster Resonance Energy Transfer Parameter Estimation," Photonics **2**(4), 1027–1042 (2015).

42. MATLAB, "MATLAB FITTING TOOLBOX," https://www.mathworks.com/help/pdf_doc/curvefit/curvefit.pdf.

43. J. T. Smith, R. Yao, N. Sinsuebphon, A. Rudkouskaya, N. Un, J. Mazurkiewicz, M. Barroso, P. Yan, and X. Intes, "Fast fit-free analysis of fluorescence lifetime imaging via deep learning," Proc. Natl. Acad. Sci. **116**(48), 24019–24030 (2019).